Probabilités : Généralités

• Une famille $\mathcal{A}$ de parties de E est une <u>tribu</u> (ou <u>σ-algèbre</u>) sur E si $E \in \mathcal{A}$, si $\bar{A}$, $\bar{B}$, $A \cup B$ (et donc $A \cap B$ et $A \setminus B$) appartiennent à $\mathcal{A}$ dès que A et B sont dans $\mathcal{A}$, et si $\bigcup_{i \in \mathbb{N}} A_i \in \mathcal{A}$ (et donc $\bigcap_{i \in \mathbb{N}} A_i \in \mathcal{A}$) dès que $A_i \in \mathcal{A}$ pour $i \in \mathbb{N}$.
Une probabilité est une application $p : \mathcal{A} \longrightarrow \mathbb{R}_+$ telle que $p(E) = 1$ et $p\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \sum_{i=0}^{\infty} p(A_i)$ pour toute famille $\{A_i\}_{i \in \mathbb{N}}$ d'événements disjoints 2 à 2.
(propriété de σ-additivité)

• <u>Proba. totales</u> : $p(A \cup B) = p(A) + p(B) - p(A \cap B)$

$$p\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i} p(A_i) - \sum_{i \neq j} p(A_i \cap A_j) + \dots + (-1)^{k+1} \sum_{\substack{i_1, \dots, i_k \\ \text{dist. 2 à 2}}} p(A_{i_1} \cap \dots \cap A_{i_k}) + \dots + (-1)^{n+1} p\left(\bigcap_{i=1}^{n} A_i\right)$$

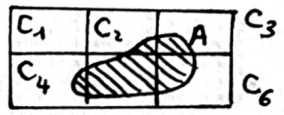
• <u>Proba. composées</u> :

$p(A/B) = \dfrac{p(A \cap B)}{p(B)}$ se généralise : $p(A_1 \cap \dots \cap A_n) = p(A_1 / A_2 \cap \dots \cap A_n) \dots p(A_{n-1} / A_n) \cdot p(A_n)$

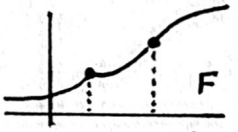
Deux événements A et B sont indépendants (resp. incompatibles) si $p(A \cap B) = p(A) \cdot p(B)$ ie $p(A/B) = p(A)$ (resp. si $p(A \cap B) = 0$)

• <u>Th. de Bayes</u> : Si $\{C_1, \dots, C_n\}$ est une partition de E et si $A \subset E$, on a :

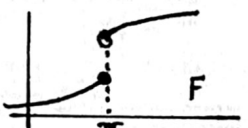
$$p(C_i / A) = \frac{p(A/C_i) \cdot p(C_i)}{\sum_{j=1}^{n} p(A/C_j) \cdot p(C_j)}$$



• <u>Variable aléatoire continue</u> : $F(x) \doteq p(X < x)$ est la <u>fonction de répartition</u> de X.
Si $F(x) = \int_{-\infty}^{x} f(t)\,dt$, f est la <u>densité de probabilité</u> de la v.a. X et X s'appelle une <u>v.a. absolument continue</u>. Notons que si f est continue par morceaux sur $\mathbb{R}$ et si $\int_{\mathbb{R}} f = 1$, alors f est une densité de probabilité d'une v.a. X dont la fonction de répartition F est dérivable en tout point sauf, peut être, aux points de discontinuité de f, et $F'(x) = f(x)$.



v.a.a. continue



v.a. mixte

• <u>Variable aléatoire mixte</u> :
Elle possède une densité f telle que $\int_{\mathbb{R}} f = 1 - a$ $(0 < a < 1)$ et des probabilités discrètes sur un ensemble dénombrable $\{x_i\}_{i \in \mathbb{N}}$ $p(X = x_i) = p_i$ et $\sum_{i \in \mathbb{N}} p_i = a$.

• La fonction de répartition $F(x) = p(X < x)$ est croissante (donc admet des limites à droite et à gauche en tout point de $\overline{\mathbb{R}}$), continue à gauche en tout point, et $F(+\infty) = 1$, $F(-\infty) = 0$.

• <u>Moments</u> : L'espérance math. (ou <u>valeur moyenne</u>) de X est, lorsqu'elle existe, l'int. de Stieltjes $E(X) = \int_{\mathbb{R}} x\,dF(x)$. Si X est discrète : $E(X) = \sum p(X = x_i)\, x_i$; si X est absolument continue : $E(X) = \int_{\mathbb{R}} x f(x)\,dx$; si φ est une fct : $E(\varphi(X)) = \int_{\mathbb{R}} \varphi(x)\,dF(x)$.
$\varphi(X) = X^n$ où $n \in \mathbb{N}^*$ permet d'obtenir (en posant $\bar{X} = E(X)$) :

$E(X^n) = \int_{\mathbb{R}} x^n\,dF(x) =$ <u>moment d'ordre n de X</u>

$E((X - \bar{X})^n) = \int_{\mathbb{R}} (x - \bar{X})^n\,dF(x) =$ moment centré d'ordre n de X.

La <u>variance</u> de X est l'espérance du carré de la v.a. centrée : $V(X) = E((X - \bar{X})^2)$.
$\sigma(X) \doteq \sqrt{V(X)} =$ <u>écart-type</u> de X.
$V(X) = E(X^2) - E(X)^2$ (form. de Koenig)

Couples aléatoires

1) Couples discrets (X,Y) : Si $p_{i,j} = p(X=x_i \text{ et } Y=y_j)$ et $p_{i,\cdot} = p(X=x_i)$, on a $p_{i,\cdot} = \sum_{j=1}^{\infty} p_{ij}$ et $p_{i,\cdot}$ est la loi marginale pour la v.a. X. X et Y sont indépendantes ssi $p_{ij} = p_{i,\cdot} \cdot p_{\cdot,j}$ $\forall i,j$

2) Cas général : $F(x,y) = p(X<x \text{ et } Y<y)$ est la fonction de répartition de (X,Y).
Si $F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(x,y)\,dx\,dy$, le couple (X,Y) est dit "absolument continu" et f est sa "densité de probabilité". Si $\Delta \subset \mathbb{R}^2$, $p(M \in \Delta) = \iint_{\Delta} f(x,y)\,dx\,dy$.

- $f(x,y) = \dfrac{\partial^2 F}{\partial x\,\partial y}$ en tt pt (x,y) où f cont. ; $\iint_{\mathbb{R}^2} f = 1$ et les lois marginales sont :

$F_1(x) = p(X<x) = \int_{-\infty}^{x}\left(\int_{\mathbb{R}} f(x,y)\,dy\right)dx$ de densité $\int_{\mathbb{R}} f(x,y)\,dy$.

$F_2(y) = p(Y<y) = \int_{-\infty}^{y}\left(\int_{\mathbb{R}} f(x,y)\,dx\right)dy$ de densité $\int_{\mathbb{R}} f(x,y)\,dx$.

← La loi du couple (X,Y) permet d'atteindre les lois marginales de X et Y.

3) X et Y sont 2 v.a indépendantes ssi les évènements $X \in I_1$ et $Y \in I_2$ sont indép. (où I_1, I_2 int. de $\mathbb{R}$), ie si $F(x,y) = F_1(x).F_2(y)$ pour tous x,y, ie $f(x,y) = f_1(x).f_2(y)$ s'il existe des densités de probabilité.
Prop. : Si X et Y sont indépendantes, alors $\varphi(X)$ et $\psi(Y)$ aussi.

4) Moments d'un couple de v.a : Si $Z = \varphi(X,Y)$, la loi de Z est donnée par la fct de répartition

$$G(z) = p(\varphi(X,Y)<z) = \begin{cases} \sum_{\varphi(x_i,y_j)<z} p(X=x_i \text{ et } Y=y_j) \text{ dans le cas discret} \\ \iint_{\varphi(x,y)<z} f(x,y)\,dx\,dy \text{ si le couple } (X,Y) \text{ possède une densité de p.} \end{cases}$$

- ex : Si $Z = X+Y$, $G(z) = \iint_{x+y<z} f(x,y)\,dx\,dy = \int_{-\infty}^{z} g(z)\,dz$ où $g(z) = \int_{\mathbb{R}} f(x,z-x)\,dx$, et si de plus X et Y sont indépendantes, $G(z) = f_1 * f_2$.
- $E(X) = \int_{\mathbb{R}} x f_1(x)\,dx = \iint_{\mathbb{R}^2} x f(x,y)\,dx\,dy$ et $E(Y) = \iint_{\mathbb{R}^2} y f(x,y)\,dx\,dy$ se calculent à partir de la loi du couple ; $E(\varphi(X,Y)) = \iint_{\mathbb{R}^2} \varphi(x,y) f(x,y)\,dx\,dy$ donne en particulier :

$\mu'_{k,n-k}(X,Y) = E(X^k, Y^{n-k}) = \iint_{\mathbb{R}^2} x^k y^{n-k} f(x,y)\,dx\,dy$ = moment d'ordre k en X et $n-k$ en Y.

$\mu_{k,n-k}(X,Y) = E((X-\bar{X})^k, (Y-\bar{Y})^{n-k})$ = moment centré d'ordre k en X et $n-k$ en Y.

- Si $\lambda \in \mathbb{R}$, $E(\lambda X+Y) = \lambda E(X) + E(Y)$, $V(\lambda) = 0$ et $V(\lambda X) = \lambda^2 V(X)$.
  Si X et Y sont indépendantes, alors $E(XY) = E(X).E(Y)$ et $V(X+Y) = V(X) + V(Y)$.

5) Coefficient de corrélation linéaire :

Notons $\mu_{2,0} = E((X-\bar{X})^2) = V(X) = \sigma^2(X)$
$\mu_{0,2} = E((Y-\bar{Y})^2) = V(Y) = \sigma^2(Y)$
$\mu_{1,1} = E((X-\bar{X})(Y-\bar{Y}))$

alors $\mu_{11}^2 \leq \mu_{20} \cdot \mu_{02}$

$$\boxed{\rho = \frac{\mu_{11}}{\sigma(X)\sigma(Y)}}$$ est le coefficient de corrélation de X et Y

On a $\rho \in [-1,1]$, et $\rho \in \{\pm 1\}$ ssi il existe une relation linéaire entre X et Y.
Si $\rho = 0$, X et Y sont "non corrélées".
Si X et Y sont indépendantes, alors elles sont non corrélées (réciproque fausse).

- On tire au sort 100 bulletins parmi 300 bulletins. Sachant que l'on a envoyé 10 bulletins à son nom parmi ces 300 bulletins, quelle est la probabilité de gagner ?

(question soulevée par Eric en juillet 92)

---

1) Envisageons d'abord le cas de 2 tirages parmi ces 300 bulletins contenant 10 bulletins gagnants :

Soient les événements : A = "j'ai gagné au 1er tirage"

B = "j'ai gagné au 2e tirage"

$$p(A \cup B) = \underbrace{p(A)}_{\frac{10}{300}} + \underbrace{p(B \cap \bar{A})}_{p(B/\bar{A}).p(\bar{A}) = \frac{10}{299} \times \frac{290}{300}}$$

soit $$p(A \cup B) = \frac{10}{300} + \frac{10}{300} \cdot \frac{290}{299}$$

2) Que l'on généralise au cas de 100 bulletins tirés au sort :

Soit A_i = " j'ai gagné au i-ème tirage " $(1 \leq i \leq 100)$

$$p = p(A_1 \cup \dots \cup A_{100})$$

$$= \underbrace{p(A_1)}_{\frac{10}{300}} + \underbrace{p(A_2 \cap \bar{A}_1)}_{\frac{10}{299} \cdot \frac{290}{300}} + \underbrace{p(A_3 \cap (\overline{A_1 \cup A_2}))}_{p(A_3/\overline{A_1 \cup A_2}).p(\overline{A_1 \cup A_2})} + \dots + \underbrace{p(A_{100} \cap (\overline{A_1 \cup \dots \cup A_{99}}))}_{p(A_{100}/\overline{A_1 \cup \dots \cup A_{99}}).p(\overline{A_1 \cup \dots \cup A_{99}})}$$

$$\frac{10}{298} \cdot \frac{C_{290}^2}{C_{300}^2} \qquad \frac{10}{201} \cdot \frac{C_{290}^{99}}{C_{300}^{99}}$$

$$= \sum_{k=0}^{99} \frac{10}{300-k} \cdot \frac{C_{290}^k}{C_{300}^k}$$

que l'on peut écrire : $$p = \frac{1}{30} \sum_{k=0}^{99} \frac{(299-k)(298-k)\dots(291-k)}{299 \cdot 298 \dots 291}$$

$$p = \frac{1}{30} \sum_{k=0}^{99} \left(1 - \frac{k}{299}\right)\left(1 - \frac{k}{298}\right) \dots \left(1 - \frac{k}{291}\right) \simeq 0{,}98$$

2ème solution :

Nombre d'événements : c'est le nbre de parties à 100 éléments d'un ensemble à 300 éléments, soit C_{300}^{100}

Nombre d'événements correspondant à un échec : c'est le nbre de parties à 100 éléments d'un ensemble à 290 éléments soit C_{290}^{100}

La probabilité d'échec est $\dfrac{C_{290}^{100}}{C_{300}^{100}} = \dfrac{290!}{190!\,190!} \cdot \dfrac{100!\,200!}{300!} = \dfrac{191.192\ldots200}{291.292\ldots300}$

La probabilité de réussite est donc

$$p = 1 - \frac{191.192\ldots200}{291.292\ldots300} \simeq 0{,}98$$

3ème solution : On cherche la probabilité d'échec.

Proba. d'échec = proba. d'échec au 1er tirage $\cap \ldots \cap$ proba. d'échec au 100ième tirage

$$= \frac{290}{300} \times \frac{289}{299} \times \ldots \times \frac{191}{201}$$

(en utilisant la formule :

$P(A \cap B) = P(A) \cdot P(B/A)$

$P(A \cap B \cap C) = P(A \cap B) \cdot P(C/A \cap B) = P(A) \cdot P(B/A) \cdot P(C/A \cap B)$

etc)

La probabilité de réussite sera $p = 1 - \dfrac{290}{300} \times \dfrac{289}{299} \times \ldots \times \dfrac{191}{201} \simeq 0{,}98$

4ème solution : Les 300 bulletins contiennent 100 bulletins gagnants et 200 bulletins perdants. Un échec signifie que nos 10 bulletins se trouvent parmi les 200 bulletins perdants : sa probabilité sera $\dfrac{C_{200}^{10}}{C_{300}^{10}} = \dfrac{290.289\ldots191}{300.299\ldots201}$. On retrouve l'expression obtenue à la 3ème solution.

Quelle est la probabilité pour que 2 personnes au moins d'un groupe de 50 personnes soient nées le même jour et le même mois ?
Exprimer cette probabilité p_n en fonction du nombre de personnes n du groupe. Tracer le graphique correspondant.

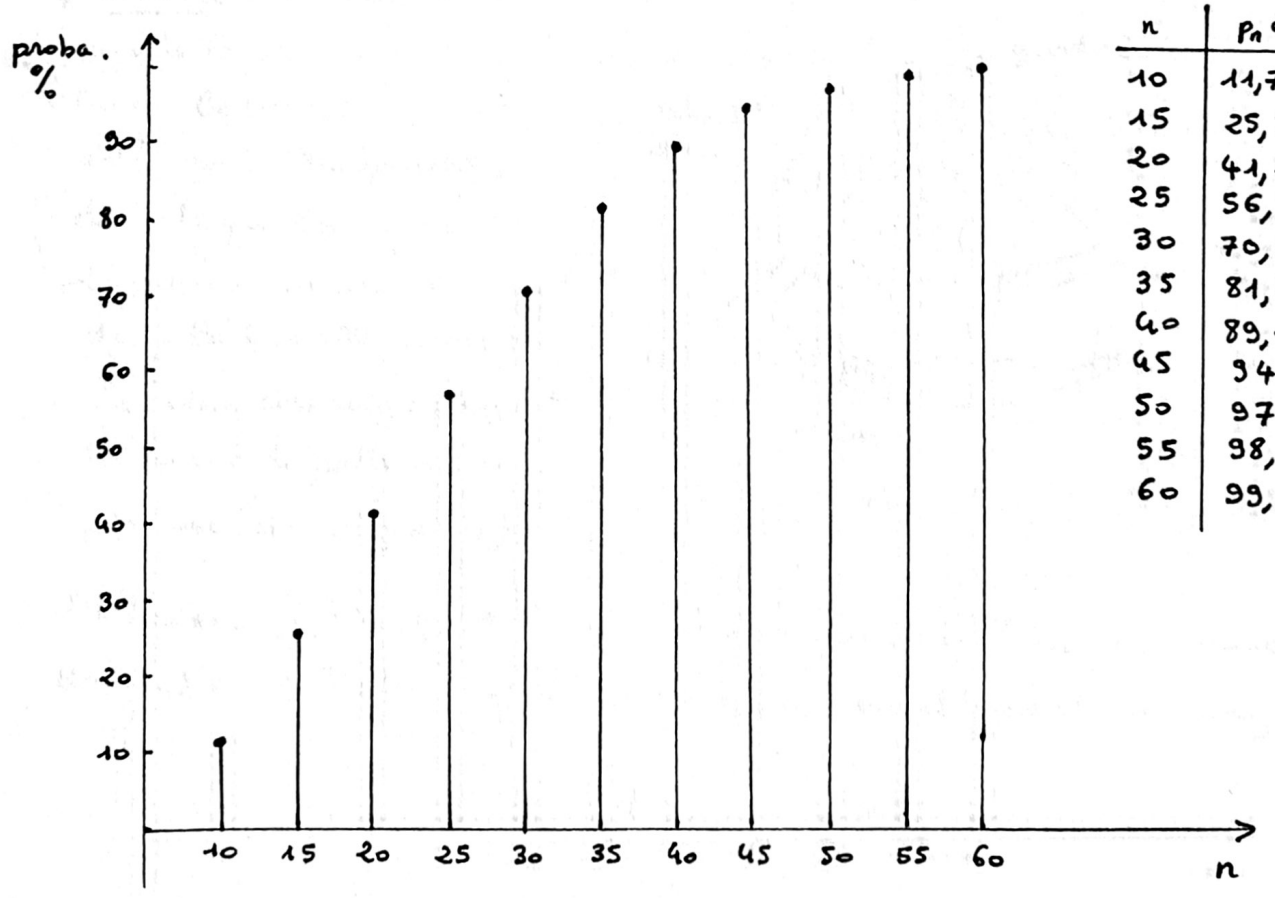
L'espace probabilisable est formé des suites $(a_1, \dots, a_{50})$ où $a_i \in \mathbb{N}_{365}$, soit 365^{50} suites possibles. La suite $(a_1, \dots, a_{50})$ décrit la situation : la 1re personne est née le a_1-ième jour de l'année, etc.
L'événement E : "2 personnes au moins sont nées le $\hat{m}$ jour et le $\hat{m}$ mois". est formé des suites $(a_1, \dots, a_{50})$ telles que 2 au moins des nbres a_i soient égaux. L'événement contraire $\bar{E}$ est formé des suites dont tous les termes sont distincts 2 à 2. Il y a $365 \times 364 \times \dots \times 316$ telles suites,

d'où $p(\bar{E}) = \frac{365}{365} \times \frac{364}{365} \times \dots \times \frac{316}{365}$ et $p(E) \simeq 97\%$

Pour un groupe de n personnes, on calcule de même :

$$p(E) = 1 - \frac{365 \cdot 364 \cdot \dots \cdot (365-n+1)}{365^n}$$

que l'on calcule facilement à l'aide d'un petit programme.



n	p_n en %
10	11,7%
15	25,3
20	41,1
25	56,9
30	70,6
35	81,4
40	89,1
45	94
50	97
55	98,6
60	99,4